APPLICATION

FOR

UNITED STATES LETTERS PATENT

TITLE: TRANSCRIPTIONAL REGULATORY FACTOR

APPLICANT: MICHAEL H. JONES

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EV399289677US

January 9, 2004

Date of Deposit

TRANSCRIPTIONAL REGULATORY FACTOR

This application is a continuation-in-part of PCT/JP99/02340, filed April 30, 1999, and claims priority from Japanese Patent Application No. 10/137631, filed April 30, 1998.

TECHNICAL FIELD

This invention relates to a novel transcriptional regulatory factor comprising bromodomains and the encoding gene.

BACKGROUND OF THE INVENTION

The bromodomain is a characteristic amino-acid motif seen in transcriptional regulatory factors and is believed to be involved in the interactions with other transcriptional regulatory factors. Proteins comprising the bromodomain, normally have one or two (Tamkun et al. (1992) Nuc. Acids Res. 20:2603; Haynes et al. (1992) Nuc. Acids Res. 20: 2603), but as many as five (Nicolas et al. (1996) Gene 175(12):233-240) bromodomain motifs. This motif has been identified in a wide range of animals, for example, in the homeotic gene (Digan et al. (1986) Dev. Biol. 114:161-169; Tamkun et al. (1992) Cell 68: 561-572) of the fruit fly (*Drosophila*), in the transcriptional regulatory genes of yeasts (Winston et al. (1987) Genetics 115:649-656; Laurent et al. (1991) Proc. Nat. Acad. Sci. USA 88:2687-2691) and in mammals (Denis et al. (1996) Genes and Devel. 10:261-271; Yang et al. (1996) Nature 382:319-324). According to a recent report (Jeanmougin et al. (1997) Trends Biochem. Sci. 22:151-153), 37 bromodomain genes, including 13 human genes are recorded in the database. In addition to the bromodomain motif of amino acid residues 59-63, the sequences adjacent to the motif are also structurally conserved, and furthermore, 4 α-helixes (Z, A, B, and C) are reported to be coded within the long 110 amino acids.

When these bromodomain-comprising transcriptional regulatory factors are compared, they all regulate signal-dependent transcription in actively proliferating cells (Tamkun et al. (1992) Cell 68:561-572; Haynes et al. (1992) Nuc. Acids Res. 20:2603). This characteristic implies that oncogenesis may occur when a gene encoding a bromodomain-containing protein undergoes abnormal regulation. In reality, six bromodomain genes have been experimentally proven to associate with oncogenesis. Three of these genes HRX/

ALL-1 (Tkachuk et al. (1992) Cell 71:691-700\; Gu et al. (1992) Cell 71:701-708); TIF1 (Miki et al. (1991) Proc. Nat. Acad. Sci. USA 88:5167-5171; Le Douarin et al. (1995) EMBO J. 14:2020-2033) and CBP (Borrow et al. (1996) Nature Genet. 14:33-41) are linked with the gene cleavage points in leukemia. All three of these proteins contain the C4HC3 (also called PHD/LAP/TRX) zinc-finger (Aasland et al. (1995) Trends Biochem. Sci. 20:56-59; Koken et al. (1995) CR Acad. Sci. III, 318:733-739; Saha et al. (1995) Proc. Nat. Acad. Sci. USA 92:9737-9741). Also, there are findings that CBP/P300 interact with p53 (Gu et al. (1997) Nature 387:819-823; Lill et al.(1997) Nature 387:823-827) and other various transcriptional factors, suggesting that CBP and the homologous gene P300 play a key-role in cancer.

The other three genes have been suggested to be linked with cancer in various ways. BRG1 interacts with retinoblastoma protein RB (Dunaief et al. (1994) Cell 79:119-130), inducing formation of flat, growth-arrested cells, and thereby showing a tumor-suppressive activity. RING3 has a homology with the fruit fly (*Drosophila*) growth control protein fsh (Haynes et al. (1989) Dev. Biol. 134:246-257) and is a serine-threonine kinase having endonuclear autophosphorylation activity. This activity has been reported to be linked to the growth phase of chronic and acute lymphocytic leukemia (Denis et al. (1996) Genes and Devel. 10:261-271). As for P/CAF, it has been reported to inhibit the interaction between E1A and p300/CBP (Yang et al. (1996) Nature 382:319-324). When P/CAF is exogenously expressed on HeLa cells, the cell cycle is inhibited. This is believed to be due to the disruption of the transcriptional regulation of E1A by the binding of P/CAF to p300/CBP. Similar to p300/CBP (Bannister and Kouzarides (1996) Nature 384:641-643), P/CAF has been reported to contain histone acetyl-transferase activity (Yang et al. (1996) Nature 382:319-324).

Thus, regulatory abnormalities of transcriptional regulatory factors comprising bromodomains are envisaged to be closely associated with various diseases, particularly, cancer and other cell-proliferation-linked diseases. Hence, attention has been focused on transcriptional regulatory factors comprising bromodomains in the recent years as novel targets for the treatment of cancer and other cell-proliferation-linked diseases.

SUMMARY OF THE INVENTION

The present invention provides a novel transcriptional regulatory factor comprising bromodomains, the encoding gene, a method of production, and a screening method for a drug-candidate compound that utilizes the protein and the gene of the present invention.

In order to solve the above-mentioned problems, EST databases were BLAST searched using various nucleotide sequences encoding known bromodomain motifs. As a result, several potential bromodomain-gene-encoding ESTs were found by the search using nucleotide sequence of *Tetrahymena thermophila* HAT A1 gene. One of these ESTs, the fetal lung cDNA library-derived EST (W17142) was found to encode an unknown gene. Therefore, isolation of full-length cDNA of EST W17142 was initiated. Specifically, primers were designed based on the EST W17142 sequence, and an amplification product was obtained by the polymerase chain reaction using testicular cDNA as the template. Then, the testicular cDNA library was screened using this amplification product as the probe, and a re-screening of the library was done using the cDNA clone comprising the above-mentioned EST sequence, thereby successfully isolating a full-length cDNA corresponding to EST W17142. By structural analysis of the protein encoded by the isolated cDNA, the present Inventors found that apart from the bromodomain, said protein had several regions and domains conserved in transcriptional regulatory factors.

Also, they found that the protein encoded by the isolated cDNA interacts with hSNF2H and hSNF2L that are implicated in the series of processes related to the chromatin-mediated transcriptional regulatory mechanism, and also with the transcription co-activator NcoA-62/Skip, which interacts with the ligand-binding domains of various nuclear receptors (VDR, RAR) and the Ski viral oncoprotein.

The transcriptional regulatory factor and the encoding gene revealed by the Inventors can be utilized for the screening of compounds inhibiting the binding between said transcriptional regulatory factor and an interacting factor, and compounds which regulate the binding activity. The compounds thus isolated are expected to be applied as pharmaceuticals.

Namely, the present invention relates to a novel transcriptional regulatory factor comprising a bromodomain and the encoding gene, as well as methods of production, and a

screening method for related-factors and drug-candidate compounds that utilize the protein and the gene of the present invention. Specifically, the present invention relates to:

1. a protein comprising the amino acid sequence of SEQ ID NO:1 or 10;

2. a transcriptional regulatory factor comprising a bromodomain and the amino acid sequence of SEQ ID NO:1 or 10, wherein one or more amino acids are replaced, deleted, added, and/or inserted;

3. a protein comprising the amino acid sequence of SEQ ID NO:1 or 10, wherein one or more amino acids are replaced, deleted, added, and/or inserted, and having an activity to bind to a protein selected from the group consisting of hSNF2H,hSNF2L,NCoA-62/Skip and homologues thereof;

4. a transcriptional regulatory factor comprising a bromodomain, and encoded by a DNA hybridizing with the DNA comprising the nucleotide sequence of SEQ ID NO:2 or 9;

5. a transcriptional regulatory factor encoded by a DNA hybridizing with the DNA comprising the nucleotide sequence of SEQ ID NO:2 or 9, and having an activity to bind to a protein selected from the group consisting of hSNF2H, hSNF2L,NCoA-62/Skip and homologues thereof;

6. a DNA encoding the transcriptional regulatory factor of any one of (1) to (5);

7. the DNA of (6), which contains the coding region of the nucleotide sequence of SEQ ID NO:2 or 9;

8. a vector containing the DNA of (6) or (7);

9. a transformant carrying, in an expressible manner, the DNA of (6) or (7);

10. a method for producing the transcriptional regulatory factor of any one of (1) to (5), the method comprising culturing the transformant of (9);

11. an antibody which binds to the transcriptional regulatory factor of any one of (1) to (5);

12. a method for screening a compound having an activity to bind to the transcriptional regulatory factor of any one of (1) to (5), the method comprising the steps of,

(a) exposing a test sample to said transcriptional regulatory factor,

(b) detecting the binding activity between the test sample and said transcriptional regulatory factor, and,

(c) selecting a compound having the binding activity to said transcriptional regulatory factor;

13. a method for screening a compound which promotes or inhibits the binding between the transcriptional regulatory factor of any one of (1) to (5) and a protein selected from the group consisting of hSNF2H, hSNF2L, NCoA-62/Skip and homologues thereof, the method comprising the steps of,

(a) exposing the transcriptional regulatory factor to hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof, in the presence of the test sample,

(b) detecting the binding activity between said transcriptional regulatory factor and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof,

(c) selecting a compound which increases or decreases said binding activity when compared with the binding activity in the absence of the test sample (control);

14. a compound which is obtainable by the method of (13), which inhibits the binding between the transcriptional regulatory factor of any one of (1) to (5) and a protein selected from the group consisting of hSNF2H, hSNF2L, NCoA-62/Skip and homologues thereof; and

15. a DNA comprising at least 15 nucleotides, which can specifically hybridize with the DNA comprising the nucleotide sequence of SEQ ID NO:2 or 9. The DNA can also be at least 351, 400, 450, 500, 700, 1000, 2200, 2500, or 3000 bp in length.

Herein, "transcriptional regulatory factor" indicates a protein that regulates gene expression. "Bromodomain" means, an amino acid motif associated with protein-protein interactions conserved within transcriptional regulatory factors linked to signal-dependent transcription.

The present invention relates to a transcriptional regulatory factor comprising a bromodomain. The amino acid sequences of the protein named "TCoA1" included in the present invention, and its variant are shown in SEQ ID NO:1 and SEQ ID NO:10, respectively, and the nucleotide sequences of their cDNA in SEQ ID NO:2 and SEQ ID NO:9, respectively (unless otherwise noted, these will be grouped as "TCoA1", hereafter). "TCoA1" is most deeply associated with the presumed proteins of nematode (*C. elegans*) chromosome III genes F26H11.2, F26H11.3a and F26H11.3b (Wilson et al. (1994) Nature 368:32-38), the function of which are unknown and which were identified by the genomic

sequence of one cosmid F26H11. When the amino acid sequence of these two proteins - the presumed nematode protein and the "TCoA1" protein - are compared, although the domain configurations are different, they are extremely alike.

Like many bromodomain proteins, "TCoA1" has one bromodomain. Being structurally similar to the TIF family, GCN5 and P/CAF, this bromodomain is situated close to the carboxyl-terminus (Jeanmougin et al. (1997) Trends Biochem. Sci. 22:151-153). Like other bromodomain proteins, "TCoA1" has a C4HC3 zinc-finger. The combination of the bromodomain and the zinc-finger has been discovered frequently in the gene cleavage points in several leukemia, so far (Tkachuk et al. (1992) Cell 71:691-700; Gu et al. (1992) Cell 71: 701-708; Miki et al. (1991) Proc. Nat. Acad. Sci. USA 88:5167-5171; Le Douarin et al. (1995) EMBO J. 14:2020-2033; Borrow et al. (1996) Nature Genet. 14:33-41). Therefore, "TCoA1" is a candidate cleavage gene associated with chromosome no. 17 q23.

"TCoA1" has numerous nuclear transport signal motifs. This indicates that "TCoA1" protein is located within the nucleus. Like other bromodomain proteins, "TCoA1" has a LXXLL motif series that likely determines the site of interaction with nuclear receptors (Heery et al. (1997) Trends Biochem. Sci. 22:151-153; Torchia et al. (1997) Nature 387:677-684). The possibility that it interacts with the receptor bound to a ligand via the LXXLL domain indicates that "TCoA1" functions as a transcriptional co-activator. In the carboxyl terminus of "TCoA1", a glutamine-rich domain is located spanning a very large region. Glutamine-rich domains have been identified in many transcriptional regulatory factors including bromodomain-containing proteins like p300/CBP (Shikama et al. (1997) Trends in Cell Biol. 2:230-236) and fsh protein of fruit fly (*Drosophila*) (Haynes et al. (1989) Dev. Biol. 134:246-257). These acidic regions have been predicted to be associated with the protein-protein interactions that determine the function as an active substance (Courey et al. (1989) Cell 59:827-836).

"TCoA1" protein has many common characteristics with other bromodomain proteins known to be linked to cell-proliferation-linked diseases such as cancer. Therefore, "TCoA1" protein may also be linked to cancer, and thus, the "TCoA1" protein and its gene, a compound that regulate the function of the "TCoA1" protein can be applied for the prevention and treatment of cancer and other cell-proliferation-linked diseases.

Moreover, the fact that hSNF2H and hSNF2L, which interact with "TCoA1", are involved in the series of processes related to the chromatin-mediated transcriptional regulatory mechanism, strongly indicates that "TCoA1" is playing some sort of a role in chromatin-mediated transcriptional regulation. Therefore, it can be conceived that "TCoA1" is playing a major role as a protein that integrates transcriptional responses towards nuclear receptors by associating with the chromatin reconstruction mechanism.

The transcriptional regulatory factor of the present invention can be prepared by methods known to one skilled in the art, as a recombinant protein made using genetic engineering techniques, and also as a natural protein. For example, a recombinant protein can be prepared by inserting DNA encoding the protein of the present invention (for example, DNA comprising the nucleotide sequence of SEQ ID NO:2 or 9) into a suitable expression vector, introducing this into a host cell, and purifying the protein from the resulting transformant. The natural protein can be acquired by preparing a column coupled with an antibody obtained by immunizing a small animal with the recombinant protein, and performing affinity chromatography for extracts of tissues or cells (for example, testis, tumor cells, etc.) expressing high levels of the transcriptional regulatory factor of the present invention.

Also, this invention features a transcriptional regulatory factor, which is functionally equivalent to the "TCoA1" protein (SEQ ID NO:1 or 10). This transcriptional regulatory factor includes, mutants of the "TCoA1" protein (SEQ ID NO:1 or 10) and "TCoA1" proteins obtained from various living organisms.

To isolate a protein functionally equivalent to a certain protein, the method of inserting a mutation into the amino acids within the protein is well known to one skilled in the art. In other words, for a person skilled in the art, the isolation of a transcriptional regulatory factor functionally equivalent to the "TCoA1" protein, is a standard procedure which can be done using, for example, the PCR-mediated, site-directed-mutation-induction system (GIBCO-BRL, Gaithersburg, Maryland), oligonucleotide-mediated, sight-directed-mutagenesis (Kramer et al. (1987) Methods in Enzymol. 154:350-367) suitably replacing amino acids that do not influence the function of the "TCoA1" protein set forth in SEQ ID NO:1 or 10. Mutations of amino acids can occur spontaneously as well. The transcriptional regulatory factor of the invention includes those comprising the amino acid sequence of

"TCoA1" protein in SEQ ID NO:1 or 10 in which one or more amino acids have been replaced, deleted, added, and/or inserted, and have a binding-activity with hSNF2H, hSNF2L and NcoA-62/Skip, and those comprising the amino acid sequence of "TCoA1" protein in SEQ ID NO:1 or 10 in which one or more amino acids have been replaced, deleted, added, and/or inserted, and comprise a bromodomain.

The term "substantially pure" as used herein in reference to a given polypeptide means that the polypeptide is substantially free from other biological macromolecules. The substantially pure polypeptide is at least 75% (e.g., at least 80, 85, 95, or 99%) pure by dry weight. Purity can be measured by any appropriate standard method, for example, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

The number of amino acids that are mutated is not particularly restricted, as long as the function of the "TCoA1" protein is maintained. Normally, it is within 50 amino acids, preferably within 30 amino acids, more preferably within 10 amino acids and even more preferably within 3 amino acids. The site of mutation may be any site, as long as the function of the "TCoA1" protein is maintained.

Proteins having amino acid sequences modified by deleting, adding and/or replacing one or more amino acid residues of a certain amino acid sequence, have been known to retain the original biological activity (Mark et al., Proc. Natl. Acad. Sci. USA (1984) 81:5662-5666; Zoller et al. Nucleic Acids Research (1982) 10:6487-6500; Wang et al., Science 224:1431-1433; Dalbadie-McFarland et al., Proc. Natl. Acad. Sci. USA (1982) 79:6409-6413).

As for the amino acid residue to be mutated, it is preferable to be mutated into a different amino acid in which the properties of the amino acid side-chain are conserved. Examples of properties of amino acid side chains are, hydrophobic amino acids (A, I, L, M, F, P, W, Y, V), hydrophilic amino acids (R, D, N, C, E, Q, G, H, K, S, T), and amino acids comprising the following side chains: an aliphatic side-chain (G, A, V, L, I, P); a hydroxyl group containing side-chain (S, T, Y); a sulfur atom containing side-chain (C, M); a carboxylic acid and amide containing side-chain (D, N, E, Q); a base containing side-chain (R, K, H); and an aromatic containing side-chain (H, F, Y, W). (The parenthetic letters indicate the one-letter codes of amino acids). A "conservative amino acid substitution is a

replacement of one amino acid belonging to one of the above groups with another amino acid in the same group.

In the present invention, the protein having several deletions in the amino acid sequence of the "TCoA1" protein (SEQ ID NO:1 or 10) includes a partial peptide comprising binding-activity with hSNF2H, hSNF2L, NcoA-62/Skip or homologues thereof. As described in Example 6 (Fig. 5), the N-terminus of the "TCoA1" protein has a binding-activity with hSNF2H, hSNF2L, NcoA-62/Skip or homologues thereof. Peptides such as these, inhibit the binding between "TCoA1" protein and the above binding-proteins *in vivo*, and thus can be used to inhibit the functions of the "TCoA1" protein *in vivo*.

A fusion protein including the "TCoA1" protein can be given as an example of a protein into which several amino acid residues have been added to the amino acid sequence of the "TCoA1" protein (SEQ ID NO:1 or 10). Fusion proteins are, fusions of the "TCoA1" protein and other peptides or proteins, and are included in the present invention. Fusion proteins can be made by techniques well known to a person skilled in the art, such as by linking the DNA encoding the "TCoA1" protein of the invention with DNA encoding other peptides or proteins, so as the frames match, inserting this into an expression vector and expressing it in a host. There is no restriction as to the peptides or proteins fused to the protein of the present invention.

Known peptides, for example, FLAG (Hopp et al., Biotechnology (1988) 6:1204-1210), 6xHis containing six His (histidine) residues, 10xHis, Influenza agglutinin (HA), human c-myc fragment, VSP-GP fragment, p18HIV fragment, T7-tag, HSV-tag, E-tag, SV40T antigen fragment, lck tag, α-tubulin fragment, B-tag, Protein C fragment, and such, can be used as peptides that are fused to the protein of the present invention. Examples of proteins that are fused to protein of the invention are, GST (glutathione-S-transferase), Influenza agglutinin (HA), immunoglobulin constant region, β-galactosidase, MBP (maltose-binding protein), and such.

Fusion proteins can be prepared by fusing commercially available DNA encoding these peptides or proteins with the DNA encoding the protein of the present invention and expressing the fused DNA prepared.

The hybridization technique (Sambrook et al., Molecular Cloning 2nd ed. 9.47-9.58, Cold Spring Harbor Lab. press, 1989) is well known to one skilled in the art as an alternative

method for isolating a protein functionally equivalent to a certain protein. In other words, for a person skilled in the art, it is a general procedure to obtain a transcriptional regulatory factor functionally equivalent to the "TCoA1" protein, by isolating DNA having a high homology with the whole or part of the DNA encoding the "TCoA1" protein of SEQ ID NO:2 using the hybridization technique. The transcriptional regulatory factor of the present invention, includes transcriptional regulatory factors comprising bromodomains which are encoded by the DNA hybridizing with the DNA encoding "TCoA1" protein of SEQ ID NO:2. Animals which can be used to isolate a functionally equivalent transcriptional regulatory factor are, apart from humans, for example, mice, rats, cattle, monkeys and pigs, but there are no restrictions to the animal used. The stringency of hybridization is defined as equilibrium hybridization under the following conditions: 42°C, 2 x SSC, 0.1% SDS (low stringency); 50°C, 2 x SSC, 0.1% SDS (medium stringency); and 65°C, 2 x SSC, 0.1% SDS (high stringency). If washings are necessary to achieve equilibrium, the washings are performed with the hybridization solution for the particular stringency desired. In general, the higher the temperature, the higher is the homology between two strands hybridizing at equilibrium. However, several factors other than temperature can influence the stringency of hybridization and one skilled in the art can suitably select the factors to accomplish a similar stringency.

In place of hybridization, the gene amplification method using a primer synthesized based on the sequence information of the DNA sequence of SEQ ID NO:9 encoding the "TCoA1" protein, for example, the polymerase chain reaction (PCR) method can be utilized to isolate a DNA encoding a transcriptional regulatory factor functionally equivalent to the "TCoA1" protein.

Proteins encoded by the DNA isolated through the above hybridization technique or gene amplification techniques, normally have a high homology to the amino acid sequence of the "TCoA1" protein. "High homology" refers to, normally a homology of 40% or higher, preferably 60% or higher, more preferably 80% or higher, even more preferably 95% or higher with the amino acid sequence of the "TCoA1" protein. The homology of a protein can be determined by following the algorithm in "Wilbur, W.J. and Lipman, D.J. Proc. Natl. Acad. Sci. USA (1983) 80, 726-730".

The "percent identity" of two amino acid sequences or of two nucleic acids is determined using the algorithm of Karlin and Altschul (Proc. Natl. Acad. Sci. USA 87:2264-2268, 1990), modified as in Karlin and Altschul (Proc. Natl. Acad. Sci. USA 90:5873-5877, 1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al. (J. Mol. Biol. 215:403-410, 1990). BLAST nucleotide searches are performed with the NBLAST program, score = 100, wordlength = 12. BLAST protein searches are performed with the XBLAST program, score = 50, wordlength = 3. Where gaps exist between two sequences, Gapped BLAST is utilized as described in Altschul et al. (Nucleic Acids Res. 25:3389-3402, 1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) are used. See http://www.ncbi.nlm.nih.gov.

An "isolated nucleic acid" is a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of different (i) DNA molecules, (ii) transfected cells, or (iii) cell clones: e.g., as these occur in a DNA library such as a cDNA or genomic DNA library.

Transcriptional regulatory factors functionally equivalent to the “TCoA1” protein (SEQ ID NO:1 or 10) isolated by the above hybridization technique or gene amplification techniques include, those having a binding activity with hSNF2H, hSNF2L and NcoA-62/Skip, and a high homology in the primary structure with the “TCoA1” protein (SEQ ID NO:1 or 10), and those having the bromodomain, which is a motif thought to be vital to the

function linked with cancer, and a high homology in the primary structure with the "TCoA1" protein (SEQ ID NO:10).

Other than the bromodomain, these transcriptional regulatory factors also comprise sequences involved in the interactions with other proteins (for example, leucine-zipper, LXXLL motif), sequences involved in the binding with DNA (for example, zinc-finger), and nuclear transport signals.

The existence of the bromodomain within a protein can be determined by searching the bromodomain motif PROSITE database on DNASIS (Hitachi Software Engineering).

This invention also relates to a DNA encoding the above transcriptional regulatory factor. There is no restriction as to the DNA of the present invention as long as it encodes the transcriptional regulatory factor of the invention, and includes cDNA, genomic DNA and chemically synthesized DNA. Also as long as they can encode the protein of the invention, DNAs comprising arbitrary sequences based on the degeneracy of the genetic code are also included. cDNA encoding the protein of the invention can be prepared, for example, by preparing a primer based on nucleotide information (for example, SEQ ID NO:9) of DNA encoding the transcriptional regulatory factor of the invention and performing plaque PCR (for example please refer, Affara NA et al. (1994) Genomics 22:205-210). In the case of genomic DNA, preparation can be done for example, by the method using commercially available "Qiagen genomic DNA kits" (Qiagen, Hilden, Germany). The nucleotide sequence of the DNA acquired can be decided by ordinary methods in the art by using, for example, the commercially available "dye terminator sequencing kit" (Applied Biosystems). The DNA of the present invention, as stated later, can be utilized for the production of a recombinant protein and gene therapy.

The present invention also features a vector into which the DNA of the present invention has been inserted. There is no restriction as to the vector to which DNA is inserted, and various vectors such as those for expressing the transcriptional regulatory factor of the present invention *in vivo* and those for preparing the recombinant protein can be used according to the objective. To express the transcriptional regulatory factor of the present invention *in vivo* (especially for gene therapy), various viral vectors and non-viral vectors can be used. Examples of viral vectors are, adenovirus vectors (pAdexLcw) and retrovirus vectors (pZIPneo), etc. Cationic liposomes can be given as examples of non-viral vectors.

Expression vectors are especially useful when using for the purpose of producing the transcriptional regulatory factor of the invention. For example, when using colibacili (*E. coli*) the "pREP4" (Qiagen, Hilden, Germany) and such vectors, when using yeast "SP-Q01" (Stratagene, La Jolla, California) and such, when using insect cells "Bac-to-Bac baculovirus expression system" (GIBCO-BRL, Gaithersburg, Maryland) are highly appropriate, but there is no restriction. Also, when using mammalian cells such as CHO cells, COS cells, NIH3T3 cells, for example, the "LacSwitch II expression system (Stratagene, La Jolla, California) is highly suitable, but there is no restriction. Insertion of the DNA of the present invention into a vector can be done using ordinary methods in the art.

The present invention also refers to a transformant, carrying, in an expressible manner, the DNA of the present invention. The transformant of the present invention includes, those carrying the above-mentioned vector into which DNA of the present invention has been inserted, and those having host genomes into which the DNA of the present invention has been integrated. As long as the DNA of the present invention is maintained in an expressible manner, no distinction is made as to the form of existence of the transformants. There is no particular restriction as to the cells into which the vector is inserted. For example, when using for the purpose of gene therapy, various cells can be used as target cells according to the type of disease. Also, when the purpose is to produce the transcriptional regulatory factor of the present invention, for example, *E. coli*, yeast, animal cells and insect cells can be used as hosts. Introduction of a vector into a cell can be done using known methods such as electroporation and calcium phosphate method.

Common methods applied in the art may be used to isolate and purify said recombinant protein from the transformant made for the production of recombinant proteins. For example, after collecting the transformant and obtaining the extracts, the objective protein can be purified and prepared by, ion exchange chromatography, reverse phase chromatography, gel filtration, or affinity chromatography where an antibody against the protein of the present invention has been immobilized in the column, or by combining several of these columns.

Also when the protein of the present invention is expressed within host cells (for example, animal cells and *E. coli*) as a fusion protein with glutathione-S-transferase protein or as a recombinant protein supplemented with multiple histidines, the expressed

recombinant protein can be purified using a glutathione column or nickel column. After purifying the fusion protein, it is also possible to exclude regions other than the objective protein by cutting with thrombin or factor-Xa as required.

The present invention also features an antibody binding to the transcriptional regulatory factor of the invention. There is no particular restriction as to the form of the antibody of the present invention and include, apart from polyclonal antibodies, monoclonal antibodies as well. The antiserum obtained by immunizing animals such as rabbits with the transcriptional regulatory factor of the present invention, polyclonal and monoclonal antibodies of all classes, humanized antibodies made by genetic engineering, human antibodies, are also included. The antibodies of the present invention can be prepared by the following methods. Polyclonal antibodies can be made by, obtaining the serum of small animals such as rabbits immunized with the transcriptional regulatory factor of the present invention, attaining a fraction recognizing only the transcriptional regulatory factor of the invention by an affinity column coupled with the protein of the present invention, and purifying immunoglobulin G or M from this fraction by a protein G or protein A column. Monoclonal antibodies can be made by immunizing small animals such as mice with the transcriptional regulatory factor of the present invention, excising the spleen from the animal, homogenizing the organ into cells, fusing the cells with mouse myeloma cells using a reagent such as polyethylene glycol, selecting clones that produce antibodies against the transcriptional regulatory factor of the invention from the fused cells (hybridomas), transplanting the obtained hybridomas into the abdominal cavity of a mouse, and extracting ascites. The obtained monoclonal antibodies can be purified by, for example, ammonium sulfate precipitation, protein A or protein G column, DEAE ion exchange chromatography, or an affinity column to which the transcriptional regulatory factor of the present invention is coupled. The antibody of the invention can be used for purifying and detecting the transcriptional regulatory factor of the invention. It can also be used as a pharmaceutical drug to inhibit the function of the present transcriptional regulatory factor. When using the antibody as a drug, in the view-point of immunogenicity, human antibodies or humanized antibodies are effective. The human antibodies or humanized antibodies can be prepared by methods commonly known to one skilled in the art. For example, human antibodies can be made by, immunizing a mouse whose immune system has been changed to that of humans,

with the transcriptional regulatory factor of the invention. Also, humanized antibodies can be prepared by, for example, cloning the antibody gene from monoclonal antibody producing cells and using the CDR graft method which transplants the antigen-recognition site of the gene into a known human antibody.

The present invention also relates to a method for screening a compound that binds to the transcriptional regulatory factor of the invention. The screening method of the invention includes the steps of, (a) exposing a test sample to the transcriptional regulatory factor of the invention, (b) detecting the binding activity between the test sample and the transcriptional regulatory factor of the invention, and (c) selecting a compound having an activity to bind to the transcriptional regulatory factor of the invention. Any test sample can be used for the screening without particular restrictions. Examples are, cell extracts, culture supernatants, synthetic low molecular weight compound libraries, purified proteins, expression products of gene libraries, synthetic peptide libraries, and so on.

Isolation of a compound that binds to the transcriptional regulatory factor using said transcriptional regulatory factor can be done using methods commonly known to one skilled in the art. The screening of a protein which binds to the transcriptional regulatory factor of the invention can be done by, for example, creating a cDNA library from tissues or cells (for example, testis tissue cells and tumor cell lines) expected to express a protein binding to the transcriptional regulatory factor of the invention using a phage vector (λgt11 and Zap, etc.), expressing this cDNA library on LB-agarose, fixing the expressed proteins on the filter, biotin-labeling the transcriptional regulatory factor of the invention or purifying it as a fusion protein with GST protein, reacting this with the above-described filter, and detecting plaques expressing the binding proteins using streptavidin or anti-GST antibody (West Western Blotting method) (Skolnik et al. (1991) Cloning of PI3 kinase-associated p85 utilizing a novel method for expression/cloning of target proteins for receptor tyrosine kinases, Cell 65:83-90). The screening of a protein binding to the transcriptional regulatory factor of the invention or its gene, can also be done by following "the two-hybrid system" ("MATCHMAKER Two-hybrid System", "Mammalian MATCHMAKER Two-Hybrid Assay Kit", "MATCHMAKER One-Hybrid System" (Clontech), "HybriZAP Two-Hybrid Vector System" (Stratagene), or Reference - "Dalton S, and Treisman R (1992) Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum

response element. Cell 68, 597-612"). In the two-hybrid system, the transcriptional regulatory factor of the invention is fused to the SRF-binding region or GAL4-binding region and expressed in yeast cells. A cDNA library, is prepared from cells expected to express a protein binding to the transcriptional regulatory factor of the invention, in a way that the library is expressed in the form of being fused to the VP16 or GAL4 transcriptional activation region. The cDNA library is then introduced into the above yeast cells and the cDNA derived from the library is isolated from the positive clones detected (when a protein binding to the transcriptional regulatory factor of the invention is expressed in yeast cells, the binding of the two activates a reporter gene making positive clones detectable). A protein binding to the transcriptional regulatory factor of the invention can be recovered by, introducing the cDNA isolated above to *E.coli* and expressing the protein encoded by said cDNA.

Also, a protein binding to the transcriptional regulatory factor of the invention can be screened by, applying the culture supernatants or cell extracts of cells expected to express a protein binding to the transcriptional regulatory factor of the invention onto an affinity column in which the protein of the invention is immobilized and purifying the protein that binds specifically to the column.

The method of screening molecules that bind when the immobilized transcriptional regulatory factor of the invention is exposed to synthetic chemical compounds, or natural substance banks, or a random phage peptide display library, or the method of screening using high-throughput based on combinatorial chemistry techniques (Wrighton et al., Small peptides as potent mimetics of the protein hormone erythropoietin, Science (UNITED STATES) (1996), 273:458-464; Verdine G. L., The combinatorial chemistry of nature, Nature (ENGLAND) (1996) 384:11-13; Hogan J.C., Jr., Directed combinatorial chemistry. Nature (ENGLAND) (1996) 384:17-19) to isolate low molecular weight compounds, proteins (or their genes) and peptides are methods well known to one skilled in the art.

A biosensor using the surface plasmon resonance phenomenon may be used as a mean for detecting or quantifying the bound compound in the present invention. When such a biosensor is used, the interaction between the protein of the invention and a test compound can be observed real-time as a surface plasmon resonance signal, using only a minute amount of proteins without labeling (for example, BIAcore, Pharmacia). Therefore, it is possible to

evaluate the binding between the transcriptional regulatory factor of the invention and a test compound using a biosensor such as BIAcore.

The present invention also relates to a method for screening a compound able to promote or inhibit the binding between the transcriptional regulatory factor of the invention and an interacting-protein. Detection of a binding between the TCoA1 protein and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof enabled such a screening. This screening can be done using the method comprising the steps of: (a) exposing the transcriptional regulatory factor of the invention to hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof, under the presence of a test sample; (b) detecting the binding activity between the transcriptional regulatory factor of the invention and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof; and (c) selecting a compound which decreases said binding-activity when compared with the assay in the absence of a test sample (control).

There are no particular restrictions as to the test sample used. Examples are, cell extracts, culture supernatants, libraries of synthetic low molecular weight compounds, purified proteins, expression products of gene libraries, synthetic peptide libraries, etc. The compound isolated by the above-described screening of a protein binding to the protein of the invention may also be used as a test sample.

The transcriptional regulatory factor of the invention used for the screening may be a whole protein or a partial peptide comprising binding regions with hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof. hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof used for the screening may be whole proteins or partial peptides comprising binding regions with the transcriptional regulatory factor of the invention.

The detection of the binding activity between the transcriptional regulatory factor of the invention and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof, can be performed, for example, as follows.

A test sample and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof is added to the transcriptional regulatory factor of the invention immobilized on a microplate, reacted with a mouse or rabbit antibody against hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof, further reacted with an anti-mouse or anti-rabbit antibody labeled with peroxidase, alkaline phosphatase and such, a labeled enzyme substrate is added and the enzyme activity is measured. Compounds that show an enzyme activity that is lower to or higher than that in

the absence of a test sample, are selected. Thereby, compounds having an activity to promote or inhibit the binding between the transcriptional regulatory factor of the invention and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof are obtained.

This screening may be performed also by, using hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof as the immobilized protein, and the transcriptional regulatory factor of the invention as the protein that is added with the test sample.

Also, the transcriptional regulatory factor of the invention or hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof added together with the test sample may be directly labeled with peroxidase, or alkaline phosphatase, or used as a fusion protein with such enzymes. Compounds having an activity that activates or inhibits the binding between the transcriptional regulatory factor of the invention and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof may also be selected by, expressing as fusion proteins with enzymes other than the above, such as, luciferase, β-galactosidase, or GFP protein and measuring the inhibition or promotion of the enzyme activity by a test sample.

The mammalian two-hybrid system (Clontech, Palo Alto) can also be used to screen a compound that promotes or inhibits the binding between the transcriptional regulatory factor of the invention and an interacting-protein. Namely, using the two-hybrid system, the transcriptional regulatory factor of the invention and an interacting-protein is expressed in mammalian cells, a test sample is added to said mammalian cells, and then reporter-activity is measured. The detected reporter-activity is compared, and compounds that give a value that is lower to or higher than the reporter-activity in the absence of a test sample, are selected. Thus, a compound that promotes or inhibits the binding between the transcriptional regulatory factor of the invention and hSNF2H, hSNF2L, NCoA-62/Skip or homologues thereof can be obtained.

A compound screened by the screening of the invention may be applied for the prevention and treatment of cancer and other cell-proliferation-linked diseases. When using the isolated compound as a pharmaceutical for humans and other mammals, such as, mice, rats, guinea-pigs, rabbits, chicken, cats, dogs, sheep, pigs, monkeys, baboons, chimpanzees, the isolated compound can be directly administered or can be formulated into a dosage form using known pharmaceutical preparation methods. For example, according to the need, the drugs can be taken orally as sugar-coated tablets, capsules, elixirs and microcapsules or non-

orally in the form of injections of sterile solutions or suspensions with water or any other pharmaceutically acceptable liquid. For example, the compounds can be mixed with pharmacologically acceptable carriers or medium, specifically, sterilized water, physiological saline, plant-oil, emulsifiers, solvents, surfactants, stabilizers, flavoring agents, excipients, vehicles, preservatives and binders, in a unit dose form required for generally accepted drug implementation. The amount of active ingredients in these preparations makes a suitable dosage within the indicated range acquirable.

Examples for additives which can be mixed to tablets and capsules are, binders such as gelatin, corn starch, tragacanth gum and arabic gum; excipients such as crystalline cellulose; swelling agents such as corn starch, gelatin and alginic acid; lubricants such as magnesium stearate; sweeteners such as sucrose, lactose or saccharin; flavoring agents such as peppermint, Gaultheria adenothrix oil and cherry. When the unit dosage form is a capsule, a liquid carrier, such as oil, can also be included in the above ingredients. Sterile composites for injections can be formulated following normal drug implementations using vehicles such as distilled water used for injections.

Physiological saline, glucose, and other isotonic liquids including adjuvants, such as D-sorbitol, D-mannose, D-mannitol, and sodium chloride, can be used as aqueous solutions for injections. These can be used in conjunction with suitable solubilizers, such as alcohol, specifically ethanol, polyalcohols such as propylene glycol and polyethylene glycol, non-ionic surfactants, such as Polysorbate 80 (TM) and HCO-50.

Sesame oil or Soy-bean oil can be used as a oleaginous liquid and may be used in conjunction with benzyl benzoate or benzyl alcohol as a solubilizers; may be formulated with a buffer such as phosphate buffer and sodium acetate buffer; a pain-killer such as procaine hydrochloride; a stabilizer such as benzyl alcohol, phenol; and an anti-oxidant. The prepared injection is filled into a suitable ampule.

Methods well known to one skilled in the art may be used to administer a pharmaceutical compound to patients, for example as intraarterial, intravenous, percutaneous injections and also as intranasal, transbronchial, intramuscular or oral administrations. The dosage and method of administration vary according to the body-weight and age of a patient and the administration method but one skilled in the art can suitably select them. If said compound is encodable by a DNA, said DNA can be inserted into a vector for gene therapy

and perform the therapy. The dosage and method of administration vary according to the body-weight, age, and symptoms of a patient but one skilled in the art can select them suitably.

For example, although there are some differences according to the symptoms, the dose of a compound that binds with the transcriptional regulatory factor of the present invention and regulates its activity is about 0.1 mg to about 100 mg per day, preferably about 1.0 mg to about 50 mg per day and more preferably about 1.0 mg to about 20 mg per day, when administered orally to a normal adult (weight 60 kg).

When administering parenterally in the form of an injection to a normal adult (weight 60 kg), although there are some differences according to the patient, target organ, symptoms and method of administration, it is convenient to intravenously inject a dose of about 0.01 mg to about 30 mg per day, preferably about 0.1 to about 20 mg per day and more preferably about 0.1 to about 10 mg per day. Also, in the case of other animals too, it is possible to administer an amount converted to 60 kg of body-weight.

This invention also features a DNA containing at least 15 nucleotides, which can specifically hybridize with DNA encoding the "TCoA1" protein. The term "specifically hybridize" as used herein, indicates that cross-hybridization does not occur significantly with DNA encoding other proteins, in the above-mentioned hybridizing conditions, preferably under stringent hybridizing conditions. Such DNA includes, probes, primers, nucleotides and nucleotide derivatives (for example, antisense oligonucleotides and ribozymes), which specifically hybridize with DNA encoding the protein of the invention or its complementary DNA.

The present invention includes an antisense oligonucleotide that hybridizes with any site within the nucleotide sequence of SEQ ID NO:2 or 9. This antisense oligonucleotide is preferably that against the at least 15 continuous nucleotides in the nucleotide sequence of SEQ ID NO:2 or 9. The above-mentioned antisense oligonucleotide, which contains an initiation codon in the above-mentioned at least 15 continuous nucleotides, is even more preferred.

Derivatives or modified products of antisense oligonucleotides can be used as antisense oligonucleotides. Examples of such modified products are, lower alkyl

phosphonate modifications such as methyl-phosphonate-type or ethyl-phosphonate-type, phosphothioate modifications and phosphoramidate modifications.

The term “antisense oligonucleotides” as used herein means, not only those in which the entire nucleotides corresponding to those constituting a specified region of a DNA or mRNA are complementary, but also those having a mismatch of one or more nucleotides, as long as DNA or mRNA and an oligonucleotide can specifically hybridize with the nucleotide sequence of SEQ ID NO:9.

Such DNAs are indicated as those having, in the “at least 15 continuous nucleotide sequence region”, a homology of at least 70% or higher, preferably at 80% or higher, more preferably 90% or higher, even more preferably 95% or higher. The algorithm stated herein can be used to determine homology. Such DNAs are useful as probes for the isolation or detection of DNA encoding the protein of the invention as stated in a later example or as a primer used for amplifications.

The antisense oligonucleotide derivative of the present invention, acts upon cells producing the protein of the invention by binding to the DNA or mRNA encoding the protein and inhibits its transcription or translation, promotes the degradation of the mRNA, inhibiting the expression of the protein of the invention resulting in the inhibition of the protein’s function.

The antisense oligonucleotide derivative of the present invention can be made into an external preparation such as a liniment and a poultice by mixing with a suitable base material, which is inactive against the derivatives.

Also, as needed, the derivatives can be formulated into tablets, powders, granules, capsules, liposome capsules, injections, solutions, nose-drops and freeze-drying agents by adding excipients, isotonic agents, solubilizers, stabilizers, preservatives, pain-killers, and such. These can be prepared by following usual methods.

The antisense oligonucleotide derivative is given to the patient by, directly applying onto the ailing site or by injecting into a blood vessel so that it will reach the site of ailment. An antisense-mounting medium can also be used to increase durability and membrane-permeability. Examples are, liposome, poly-L-lysine, lipid, cholesterol, lipofectin or derivatives of these.

The dosage of the antisense oligonucleotide derivative of the present invention can be adjusted suitably according to the patient's condition and used in desired amounts. For example, a dose range of 0.1 to 100 mg/kg, preferably 0.1 to 50 mg/kg can be administered.

The antisense oligonucleotide of the invention inhibits the expression of the protein of the invention and thereby useful for suppressing the biological activity of the protein of the invention. Also, expression-inhibitors comprising the antisense oligonucleotide of the invention are useful in the point that they can inhibit the biological activity of the protein of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the alignment of the domains identified in "TCoA1". The symbols within the figure are shown below.

CH4C3: CH4C3 zinc-finger; bHLH: basic helix-loop-helix; Q-rich: glutamine-rich; C2HC4: C2HC4 zinc-finger; BDM: bromodomain; ↑: LXXLL motif.

Fig. 2A shows the results of analysis of mono-chromosome hybrid cell panel against chromosome no. 17 using nb15G and nb15H primers. Shows the human chromosome including each hybrid. The product of 133 bp was specifically amplified in GM10498 cell-system, which is a mono-chromosome of human chromosome no. 17.

Fig. 2B shows the result of GeneBridge 4 radiation hybrid panel analysis by which the location of "TCoA1" was determined on chromosome no. 17.

Fig. 3 shows the electrophoretic pattern of the "TCoA1" expression in normal human tissues as detected by northern-blot analysis. "TCoA1" was used as the probe when hybridizing the filter in "A", and actin was used in "B". The right side of the figure shows markers.

Fig. 4 shows the results obtained by using the mammalian two-hybrid analysis system detecting the interaction between "TCoA1" and hSNF2H, hSNF2L, and NcoA-62/Skip.

Fig. 5 shows the map of the interaction between the C-terminus of TCoA1 and hSNF2H, hSNF2L, or NcoA-62/Skip.

Fig. 6 shows the map of the interaction between TCoA1 and NcoA-62/Skip. The minimal interacting region (position 224-317) is shown at the bottom.

Fig. 7 shows the map of the interaction between TCoA1 and hSNF2H. The minimal interacting region (position 921-1017) is shown at the bottom.

Fig. 8 shows the proteins that associate in the interaction with TCoA1. Unverified interactions are shown in dashed lines.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will be explained in detail below with reference to examples, but it is not construed as being limited thereto.

Example 1 Isolation of the "TCoA1" gene

(1) Identification of a novel gene comprising a bromodomain

EST database was BLAST searched using various nucleotide sequences encoding known bromodomain motifs. As a result, several potential bromodomain-gene-encoding ESTs were identified by the search using nucleotide sequence of *Tetrahymena thermophila* HAT A1 gene (Brownwell et al., (1996) Cell 84:843-851). One of these ESTs, the fetal lung cDNA library-derived EST (W17142), was discovered to provide a novel gene.

(2) Isolation of full-length nucleotide sequence

The cloning of full-length cDNA against ESTW17142 was done as follows. First the PCR primers nb15U (GGATTATGAGGGGTTGAAGAGGG/SEQ ID NO:3) and nb15L (AAGGCAACAGAGTCTGTAGCCCAA/ SEQ ID NO:4) were designed and a 119 bp amplification product was obtained by the polymerase chain reaction using testicular cDNA as the template. The amplified product was directly purified by a QIAquick (Qiagen) purifying column. Next, the testicular cDNA library (HL3024a, Clontech) was screened using this amplification product as the probe, and a re-screening of the library was done using the cDNA clone comprising the above-mentioned EST sequence. The above probe was [α-^{32}P]dCTP labeled by random priming and purified by CHROMA SPIN 10 column (Clontech). The library-filter was hybridized using ExpressHyb Hybridization Solution (Clontech) for one hour at 65°C. The filter was washed at 65°C with 0.5 x SSC. 0.1% SDS until it reached the final stringency. Next, in order to identify the hybridizing clone, autoradiography was performed at -70°C for one to three days. The same procedure was

done repeatedly until the obtained clones were linked to acquire a nucleotide sequence covering the whole coding-regions of the gene. All nucleotide sequences were determined by the ABI377 Auto Sequencer using ABI dye-terminator chemistry. Since clones of the 5' terminus were high in GC content, subcloning to the plasmid was done prior to sequence determination.

The library-screening gave 9865 bp nucleotide sequence. In this whole nucleotide sequence, an open reading frame (ORF) existed, which encoded 2993 amino acids terminating at nucleotide position 8979. This ORF was followed by 3'UTR of 877 bp until the polyA tail (Fig. 1). This sequence is believed to be the whole sequence since, the length of the sequence is comparable to the 10.5 Kb shown by northern blot analysis, and since the 5' terminus is GC rich and coincides with the existence of a CpG island seen at initiation points of many genes (Cross et al. (1995) Curr. Opin. Genet. Dev. 5:309-314). The nucleotide sequence of isolated cDNA is shown in SEQ ID NO:2 and the amino acid sequence of the protein encoded by said cDNA in SEQ ID NO:1.

(3) Determination of homology and the motif characteristics of the transcriptional factor

The motif was searched by PROSITE. The comparisons of proteins were done using Bestfit within GCG The nuclear localization signal was identified by PSORT. Motif search revealed that several conserved regions and domains were located in the amino acid sequence of presumed proteins (Fig. 1). These conserved regions had the C4HC3 zinc-finger (Aasland et al. (1995) Trends Biochem. Sci. 20:56-59; Koken et al. (1995) CR Acad. Sci. III, 318:733-739), a basic helix-loop-helix domain (Murre et al. (1989) Cell 58:537-544), an extensive hydrophobic glutamine-rich domain, CH2CH3 zinc-finger, and a bromodomain. Furthermore, there is a LXXLL motif (Torchia et al. (1997) Nature 387:677-684; Heery et al. (1997) Nature 387:733-736) that most likely furnishes the interaction with nuclear receptors. All these motifs have the characteristic to present the functions as a transcriptional regulatory factor. As a result of the PSORT program, in all, eight consensus sequences were discovered at the nuclear site, which closely associate with the above function (Robbins et al. (1991) Cell 64:615-23). Expressing the function of the gene, it was named "TCoA1" (transcriptional co-activator).

When the nucleotide sequence of "TCoA1" is analyzed upon the non-redundant DNA database, it was found that "TCoA1" has a 100% homology with 2,183 bp of the FAC1 gene (Zhu et al. (1996) Biochemica et Biophysica Acta 1309:5-8) presumed to encode a protein of 810 residues. FAC1 was initially isolated by immunoscreening of an expression library using Alz50 (Bowser et al. (1995) Dev. Neuroscience 17:20-37) monoclonal antibody. In addition to having a region that coincides spanning an extensive region with the nucleotide sequence of "TCoA1", FAC1 also coincides with the "TCoA1" results, which were obtained using the external nucleotide sequence of the region that overlaps with FAC 1, in the transcription size (Bowser et al. (1995) Dev. Neuroscience 17:20-37) and localization (Zhu et al. (1996) Biochemica et Biophysica Acta 1309:5-8). In other words, it can be envisaged that the 2673 bp nucleotide sequence of FAC1 is a partial sequence that is equivalent to the nucleotides from nucleotide position 248 of the 5' terminus' to nucleotide position 2631. Comparison of the nucleotide sequences of FCA1 and TCoA1 revealed that a single nucleotide-deleted error sequence (at position 2400 A) exists in FAC1, and thus, it can be assumed that translation terminates at an early stage together with the shift of the reading frame of ORF. Similarly, a misrecognition of the initiation point of methionine residue had been triggered by a 5' terminus sequence error in FAC1.

The predicted amino acid sequence of "TCoA1" has several extensive regions that have homologies with the presumed proteins of nematode (*C. elegans*), F26H11.2, F26H11.3a and F26H11.3b (Wilson at al., (1994) Nature 368:32-38). Results of analysis using "Gene Finder" software made the prediction of the gene that encodes these proteins possible by searching the genomic sequences contained in the F26H11 cosmid. The nucleotide sequences of "TCoA1" N terminus coincided with F26H11.g and C terminus with F26H11.I. This result showed that the both proteins presumed by "TCoA1" and FCA1 are equivalent to a single protein in the nematode, and it is believed that "TCoA1" is the human homologue of the nematode protein.

<u>Example 2</u> Chromosome mapping of "TCoA1"

To determine the chromosomal location of "TCoA1", DNA obtainable from each of the 24 monochromosomal human/rodent somatic cell lines (Dubois et al. (1993) Genomics 16:315-319) acquired from Coriell Cell Repositories, New Jersey, were amplified using the

PCR primers nb15G (CCTCAGCTGCAACAAGTCC/SEQ ID NO:5) and nb15H (GCACTGCTTTGCTGAATTTGGA/SEQ ID NO:6). As predicted, 133 bp PCR product was amplified from the GM 10567 cell system suggesting the possibility that the gene of the invention is located on human chromosome no.17 (Fig. 2A).

The "TCoA1" region locus was determined using Genebridge4 radiation hybrid panel of 91 hybrids (Walter et al. (1994) Nature Genetics 7:22-28). Screening was done by re-using primer–G and primer-H and performing PCR for that hybrid panel. By evaluating the respective hybrids as being positive or negative in regard to amplification, the binary code produced was compared with the similarity code for the marker that forms the framework map using the server at the web-address http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl to determine the chromosomal location of the gene of the invention. "TCoA1" recognized to be located in the marker D17S1557 (Fig. 2B). Only a score below 11 showing the possibility of "TCoA1" existing at a site away from D17S1557 was detected. This site coincides with the results by FISH showing FAC1 is on chromosome no. 17 q24 (Bowser (1996) Genomics 38:455-457).

To find out a more precise location of "TCoA1", screening by hierarchical PCR (Jones et al. (1994) Genomics 24:266-275) using the CEPH mega-YAC library and primers nb15S (AAGATGTTGTCTTGGAGCCGT/SEQ ID NO:7) and nb15Q (TTTTTACCATTTGCTTCAGTCCC/SEQ ID NO:8). The single clone 983d12 was identified but no information of this clone was obtainable even by searching the map information of YAC 983d12 using CEPH infoclone database (www.cephb.fr/infoclone.html). However, hybridization of Alu-PCR products showed that the two clones (902c10 and938f7) which partially overlap with 983d12, were both positive against D17S789 at the end of D17S1557. This coincides with the results of radiation hybrid obtained by the Inventors and from a cytogenetic point-of-view, means that "TCoA1" is located on chromosome no. 17 q23 (Collins et al. (1996) Proc. Natl. Acad. Sci. USA 93:14771-14775). Though slightly different, this location is close to the chromosome no. 17 q24 (Bowser (1996) Genomics 38:455-457) location reported for FAC1.

Example 3 Analysis of "TCoA1" expression

Northern hybridization was done using 240 bp cDNA probe and 16 normal tissues as panels. The probe was [α-^{32}P]dCTP labeled by random priming and purified by CHROMA SPIN 10 column (Clontech). Hybridization for northern analysis was done using ExpressHyb Hybridization Solution (Clontech) for one hour at 65°C. The filter was washed at 65°C with 0.5 x SSC, 0.1% SDS until it reached the final stringency. Next, in order to identify hybridizing transcripts, autoradiography was performed at -70°C for one to three days. mRNA blot was purchased from Clontech, and hybridization was done using the 240 bp cDNA probe equivalent to the nucleotide position 300-450. Approximately 10.5 kb mRNA was detected in almost all tissues, and the size of the transcripts was equivalent to that of ORF identified by the nucleotide sequence, and also coincided with the reported results for FAC1 (Fig. 3).

Example 4 Determination of the full-length cDNA nucleotide sequence of TCoA1

To obtain a complete cDNA, the Inventors screened the testicular cDNA library (HL3024a, Clontech) again using the 119 bp amplification product of Example 1 (2) as the probe. Screening was done under the same conditions as Example 1 (2).

When the cDNA nucleotide sequence obtained by the above screening was read, it was a sequence of 9700 nucleotides in which an inframe stop codon existed upstream the methionine initiation codon. Thus, the obtained cDNA was revealed to be full-length. The nucleotide sequence of the isolated full-length cDNA is given in SEQ ID NO:9, and the amino acid sequence of the protein encoded by said cDNA in SEQ ID NO:10.

When the nucleotide sequence of TCoA1 was compared with FAC1 (Zhu et al. (1996) Biochemica et Biophysica Acta 1309:5-8), the following nucleotide sequences coincided almost fully: position 57-1519 of FAC-1 with position 461-1917 of TcoA1, and position 1898-2622 of FAC-1 with position 1918-2643 of TCoA1. However, the position 1520-1897 of FAC-1 does not exist in the nucleotide sequence of TCoA1. The nucleotide sequence of TCoA1 has an open reading frame (ORF) coding 2781 amino acids, whereas the nucleotide sequence of FAC-1 has an ORF equivalent to a mere 810 amino acids, which is only a small part of TCoA1 beginning with a methionine initiation codon. The amino acid sequence of TCoA1 maintains two C4HC3 zinc-fingers (amino acid position 254-295) and one

bromodomain (amino acid position 2684-2747). There is also an extensive glutamine-rich region (amino acid position 1840-2400).

Example 5 Identification of proteins interacting with the N terminal region of TCoA1

Using a CDNA clone encoding the first 482 amino acids of TcoA1 including the C4HC3 zinc finger, yeast two-hybrid cDNA library (Clontech, Palo Alto) of the mouse-testis and human-brain was screened. This yeast two-hybrid cDNA library screening was done using yeast-vector PJ69-4A (James et al. (1996) Genetics 144(4):1425-36) according to the protocol of Clontech.

As a result, hSNF2H gene (Aihara et al. (1998) Cytogenet Cell Genet 81(3-4):191-3) was isolated from the human cDNA library, and the corresponding gene was isolated from the mouse cDNA library. hSNF2L gene (Aihara et al. (1998) Cytogenet Cell Genet 81(3-4):191-3) and also, a transcriptional co-activator NCoA-62 (also known as Skip) (Baudino et al. (1998) J. Biol. Chem. 273(26):16434-41, Dahl et al. (1998) Oncogene 16(12):1579-86) were isolated from the human cDNA library.

hSNF2H/2L is the human homologue of D. melanogaster's ISWI. This ISWI protein has been discovered within the chromatin reconstruction complex and this complex has been reported to be the molecular-device that reconstructs nucleosomes upon DNA in an ATPase-dependent manner (Varga-Weisz et al. (1998) Curr. Opin. Cell Biol 10(3):346-53). Within these complexes, hSNF2H and hSNF2L acts as an ATPase subunit.

Recently, there was a report suggesting the possibility that ISWI alone has an activity to reconstruct chromosomes (Corona et al. (1999) Mol. Cell 3(2):239-45). A 50 amino acid deletion at the C terminus was found when the obtained full-length sequence of hSNF2H was compared with the sequence on the database (GenBank Accession No.AB010882) and alternative splicing is believed to be occurring.

NCoA-62/Skip is a transcriptional co-activator interacting with Ski, a viral oncoprotein and ligand-binding domains of various nuclear receptors (VDR, RAR). NCoA-62/Skip also has a homology with the fruit fly (*Drosophila*) Bx42 protein induced by ecdysone.

To verify the interactions between TCoA1 and above-mentioned proteins, analysis was done using constructs of mammalian two-hybrid system (Clontech, Palo Alto) according

to protocols of Clontech. As a result, though a specific interaction could be found between TCoA1 and hSNF2H (Fig. 4), no interaction was seen for hSNF2L and Skip. Judging by the similarity of hSNF2H and hSNF2L, the lack of hSNF2L interaction was surprising and hSNF2L was probably not expressed in this system.

Example 6 The interaction map of TCoA1

The experiment for the construction of the interaction map was done as follows, using the yeast two-hybrid system. cDNAs encoding various regions (refer Fig. 5) were cloned to pAS vector (Clontech). Also cDNAs encoding the three proteins (hSNF2H, hSNF2L and NCoA-62/Skip) used in the detection of interaction with various regions of TCA1, were cloned to the pACT vector (Clontech). A combination of these vectors were introduced to a yeast-host (PJ69-4A), and the interaction between proteins expressed within said host was detected using luciferase as the reporter.

The results revealed a region that interacts with all three proteins (hSNF2H, hSNF2L and bx42 (NCoA-62/Skip)). Namely, as seen in Fig. 5, all three proteins interacted with the 85-247 amino acids of TCoA1.

This fact revealed that C4HC3 zinc finger known to be a protein interacting site was omitted from the site interacting with these 3 proteins.

Example 7 Functional analysis of the bromodomain interacting protein

The clones (hSNF2H, hSNF2L, NCoA-62/Skip) interacting with TCoA1 identified by the yeast two-hybrid screening, encode a huge polypeptide. Accordingly, the Inventors next identified the regions within these proteins that interact with TCoA1, using the yeast two-hybrid system. Specifically, a pACT vector (Clontech) constructs (Fig. 6, Fig. 7), which contained cDNA encoding a series of partially overlapping polypeptides within NcoA-62 and hSNF2H, were prepared and introduced to yeast cells (PJ69-4A) together with the pAS vector (Clontech) containing cDNA encoding the amino acids of the 1-525 site of TCoA1 protein, and the interaction between proteins expressed within said host was detected using luciferase as the reporter.

For NcoA-62, the region of approximately 450 amino acids including the complete carboxyl terminus domain of the original clone, and the series of five deletion clones in the

said region were examined (Fig. 5). As a result, amino acids of the position 224-317 within NcoA-62 were identified as the region interacting with TCoA1.

As for hSNF2H, the three deletion clones were analyzed (Fig. 7). As a result, the region within hSNF2H that interacts with TCoA1 protein was mapped to the carboxyl terminus (position 921-1017). The other clone having the same region (position 855-1017) failed to show any interaction. This may be due to the fact that this clone makes a special secondary structure.

INDUSTRIAL APPLICABILITY

The transcriptional regulatory factor and the DNA encoding said factor can be used for the treatment of cancer and other cell-proliferation-linked diseases and also for the screening of drug-candidate compounds. Furthermore, antibodies binding to the transcriptional regulatory factor of the present invention, compounds that regulate the function of said transcriptional regulatory factor, and compounds that inhibit the interaction between said transcriptional regulatory factor and other proteins, may be utilized as therapeutic agents and preventive drugs for these diseases.

What is claimed is: